

# United States Patent [19]

Watson

[11] Patent Number: 5,629,780

[45] Date of Patent: \*May 13, 1997

[54] IMAGE DATA COMPRESSION HAVING MINIMUM PERCEPTUAL ERROR

[75] Inventor: Andrew B. Watson, Los Gatos, Calif.

[73] Assignee: The United States of America as represented by the Administrator of the National Aeronautics and Space Administration, Washington, D.C.

[\*] Notice: The term of this patent shall not extend beyond the expiration date of Pat. No. 5,426,512.

[21] Appl. No.: 358,446

[22] Filed: Dec. 19, 1994

[51] Int. Cl.<sup>6</sup> H04N 1/415

[52] U.S. Cl. 358/432; 358/426; 358/433; 358/261.3; 358/539; 382/232

[58] Field of Search 358/426, 432, 358/433, 261.3, 429, 500, 515, 539; 382/232

[56] References Cited

U.S. PATENT DOCUMENTS

4,780,761 10/1988 Daly et al. 358/133

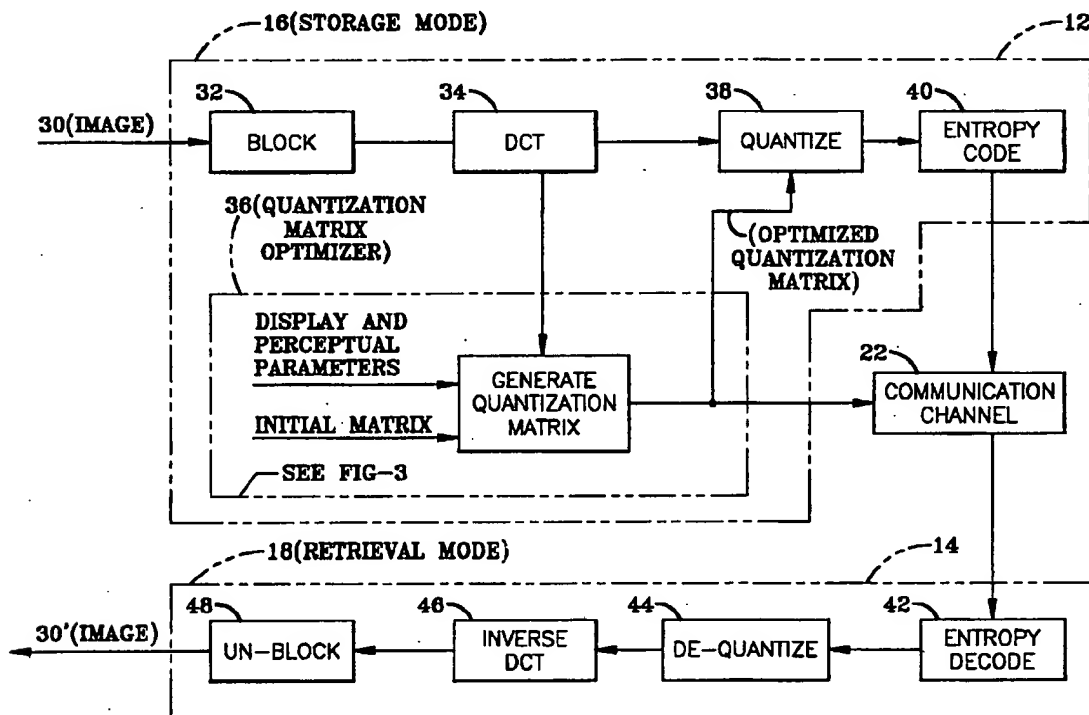
5,121,216 6/1992 Chen et al. 358/432  
5,170,264 12/1992 Saito et al. 358/432  
5,426,512 6/1995 Watson 358/432

Primary Examiner—Edward L. Coles, Sr.  
Assistant Examiner—Thomas L. Stoll  
Attorney, Agent, or Firm—Kenneth L. Warsh; Jack Mannix

## [57] ABSTRACT

A method for performing color or grayscale image compression that eliminates redundant and invisible image components. The image compression uses a Discrete Cosine Transform (DCT) and each DCT coefficient yielded by the transform is quantized by an entry in a quantization matrix which determines the perceived image quality and the bit rate of the image being compressed. The present invention adapts or customizes the quantization matrix to the image being compressed. The quantization matrix comprises visual masking by luminance and contrast techniques and by an error pooling technique all resulting in a minimum perceptual error for any given bit rate, or minimum bit rate for a given perceptual error.

10 Claims, 7 Drawing Sheets  
(1 of 7 Drawing(s) in Color)



compute average  
DC coefficient value  
(col. 8,  
claim 3).

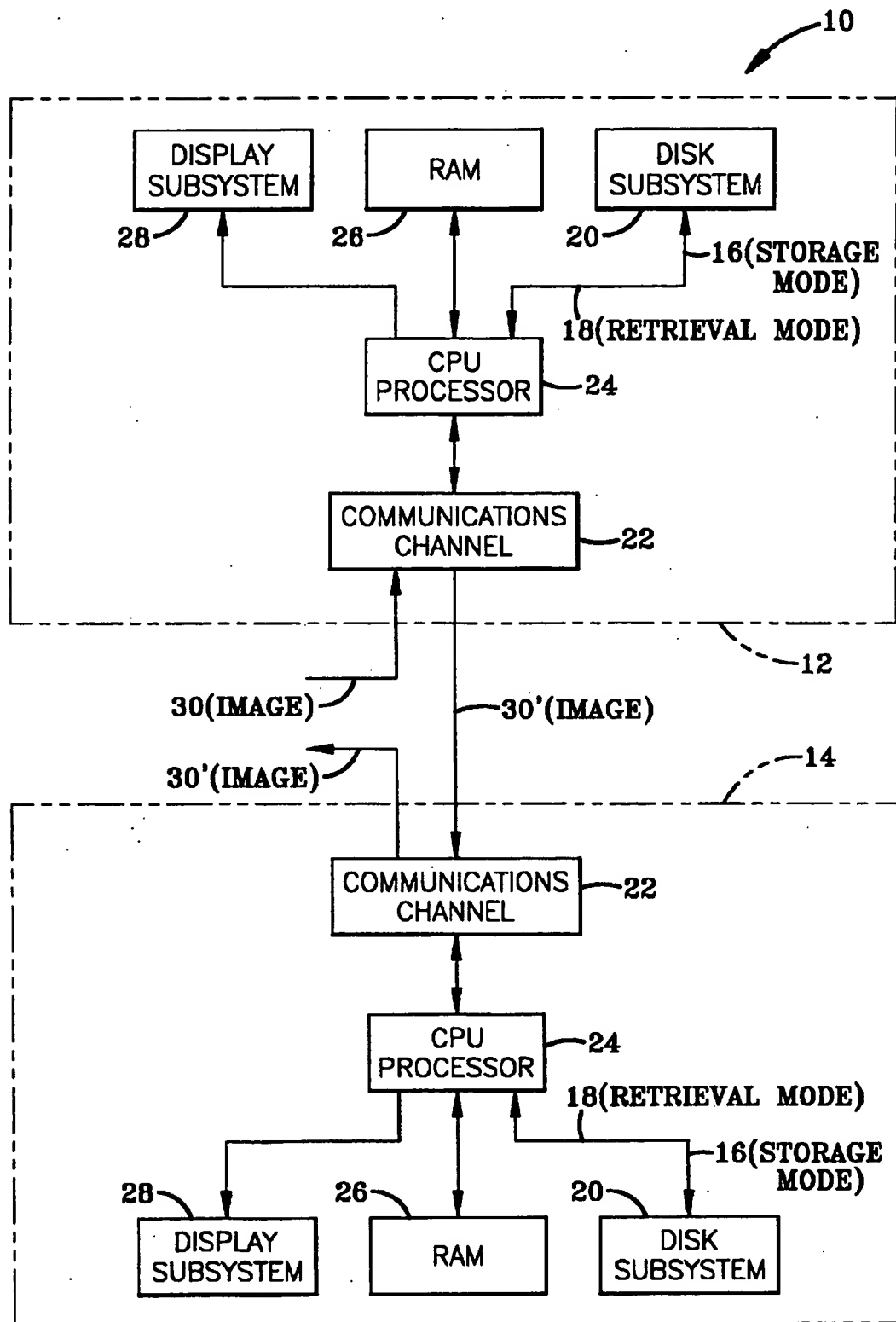


FIG-1

FIG-2

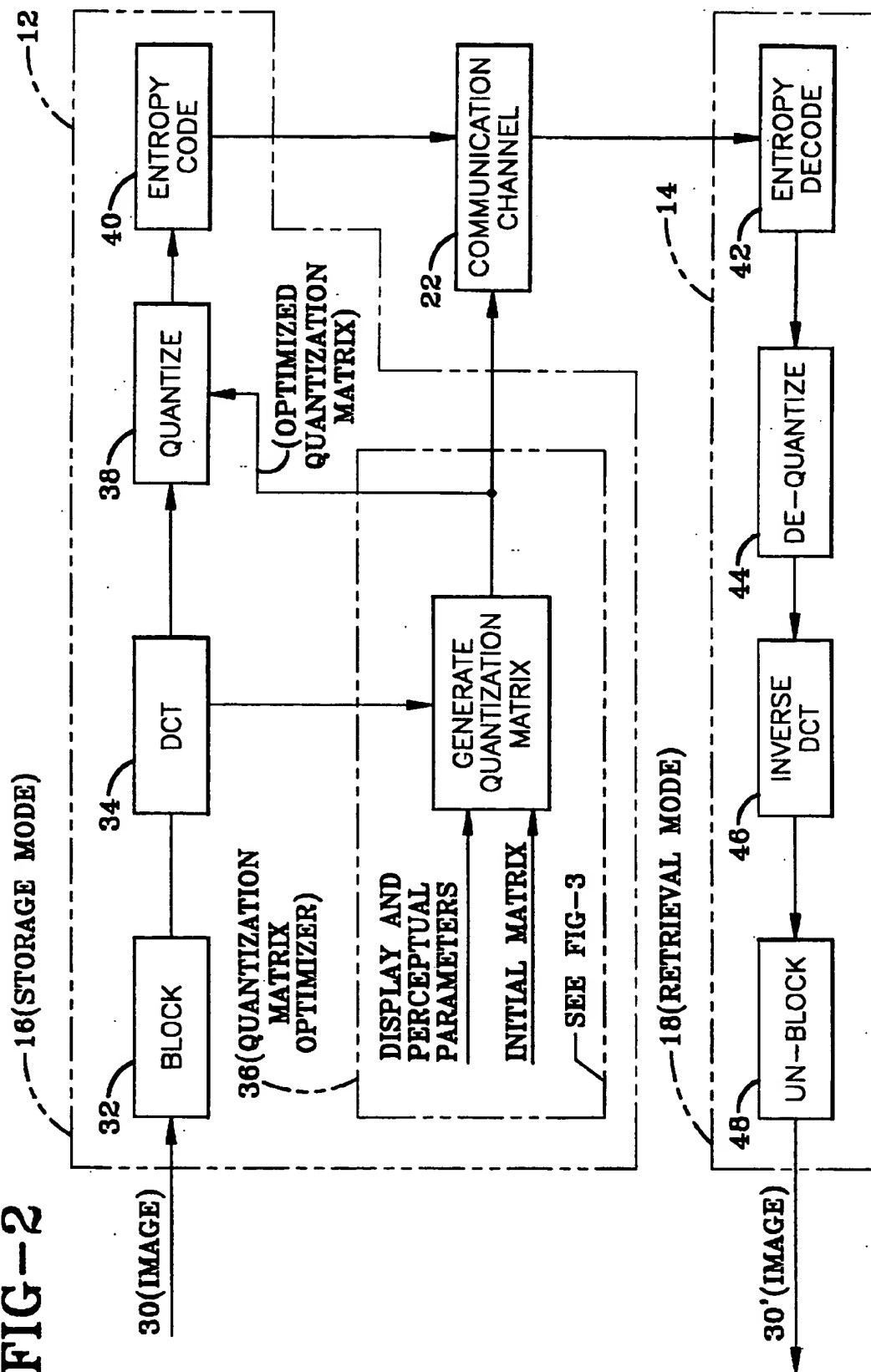
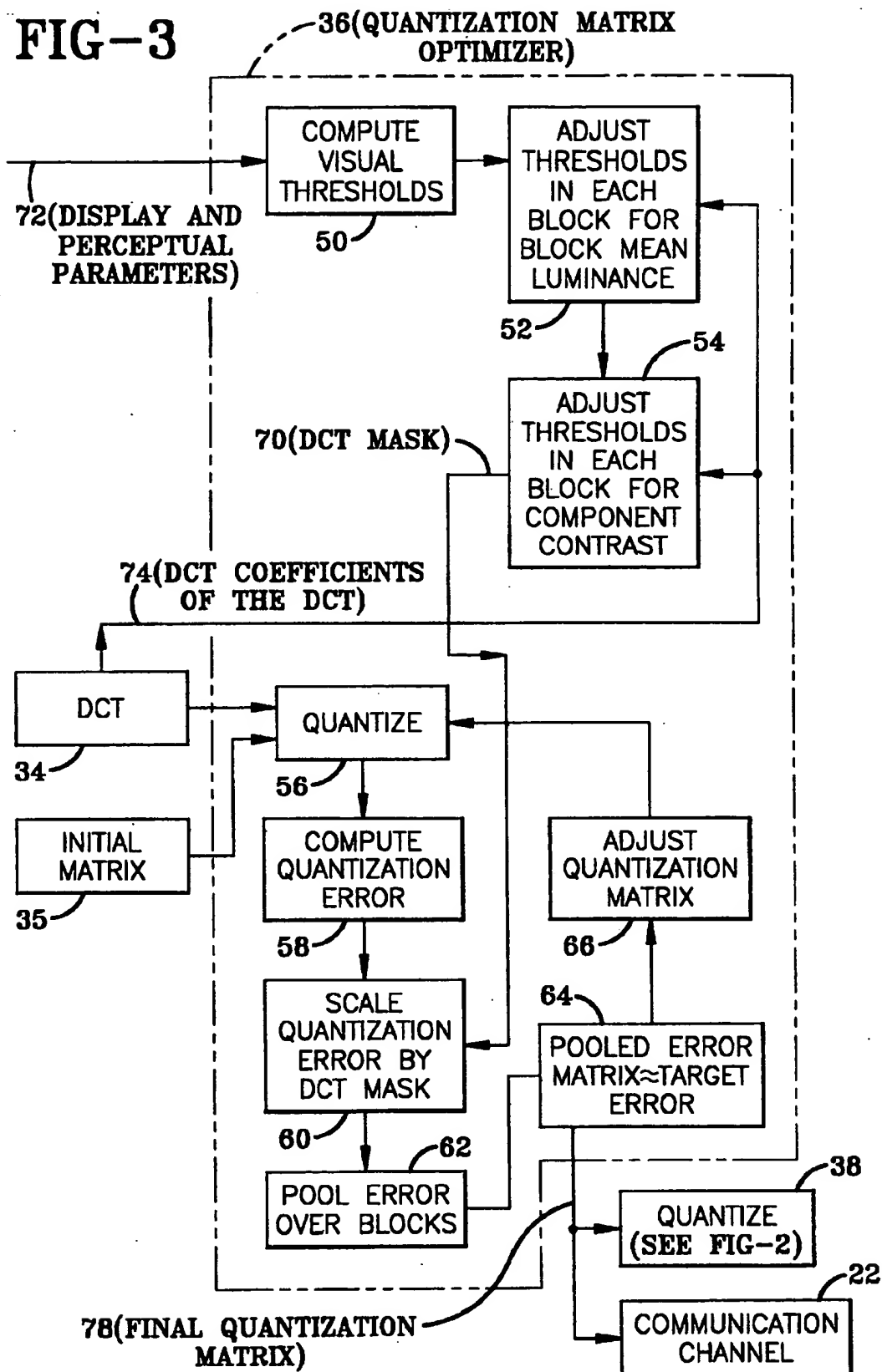


FIG-3



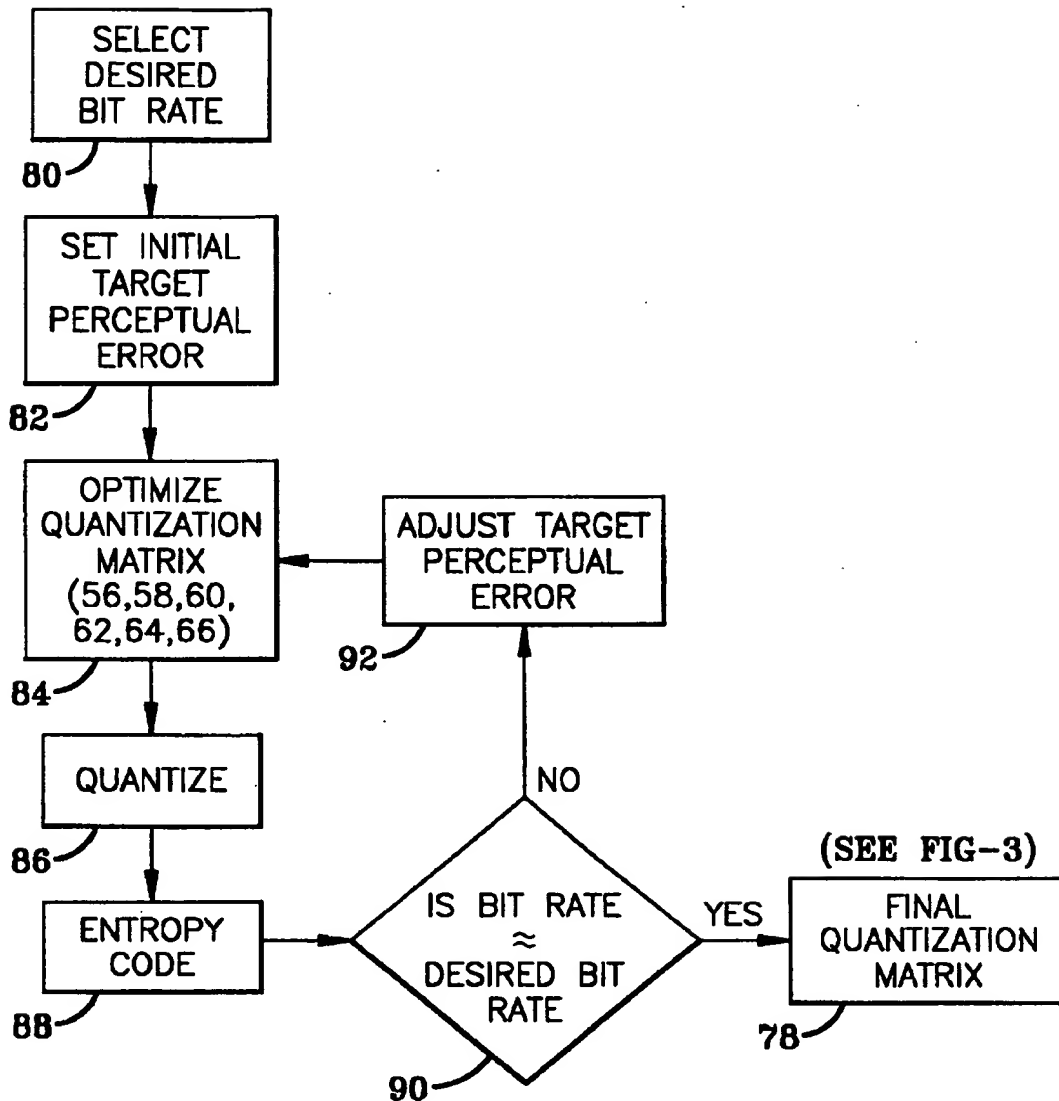
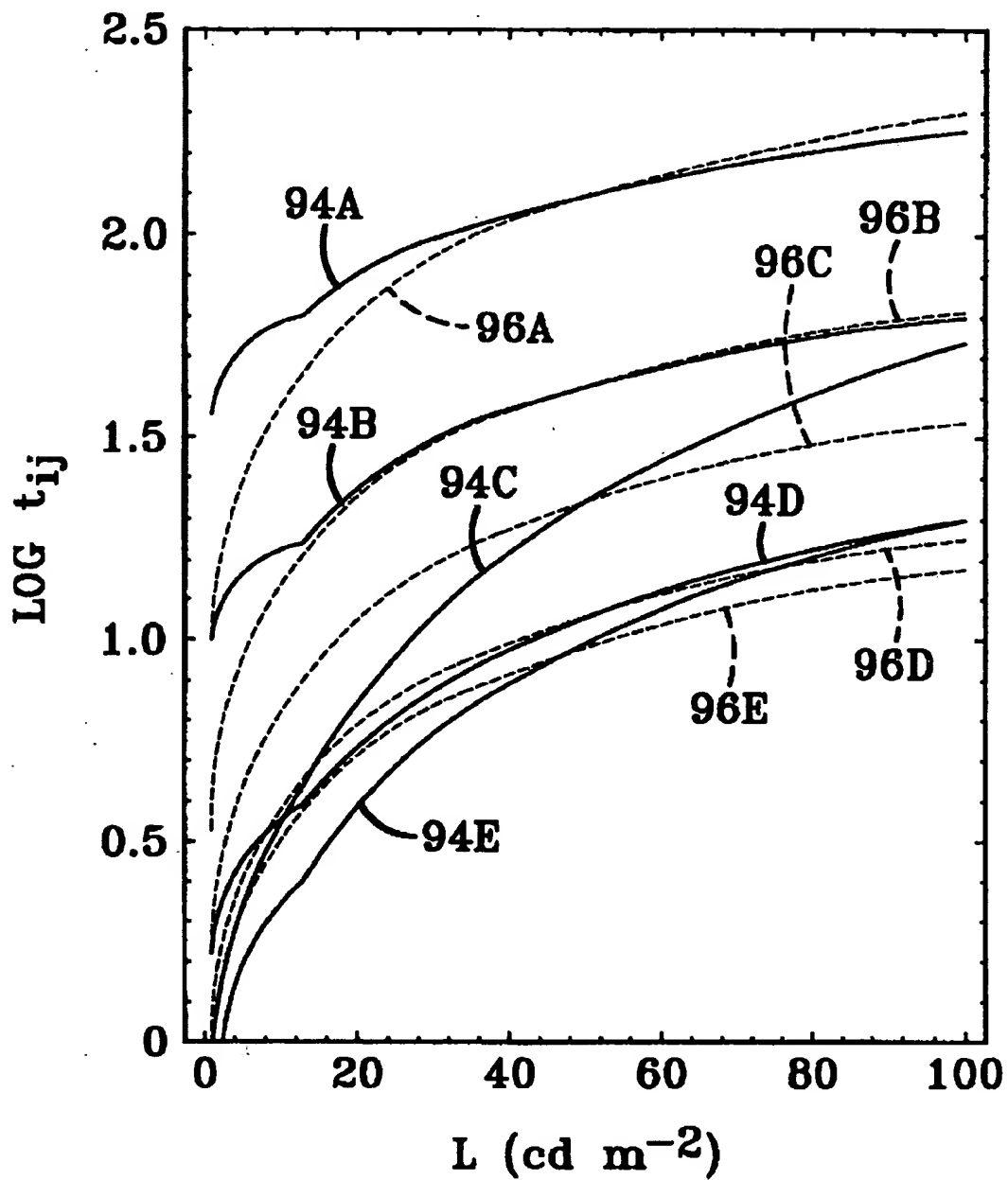


FIG-4

**FIG-5**

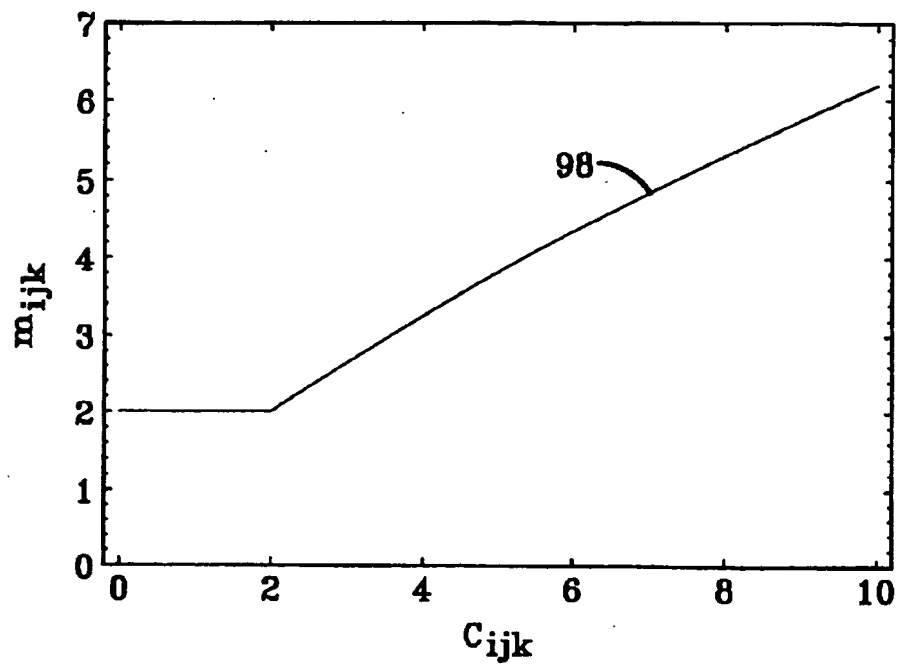


FIG-6

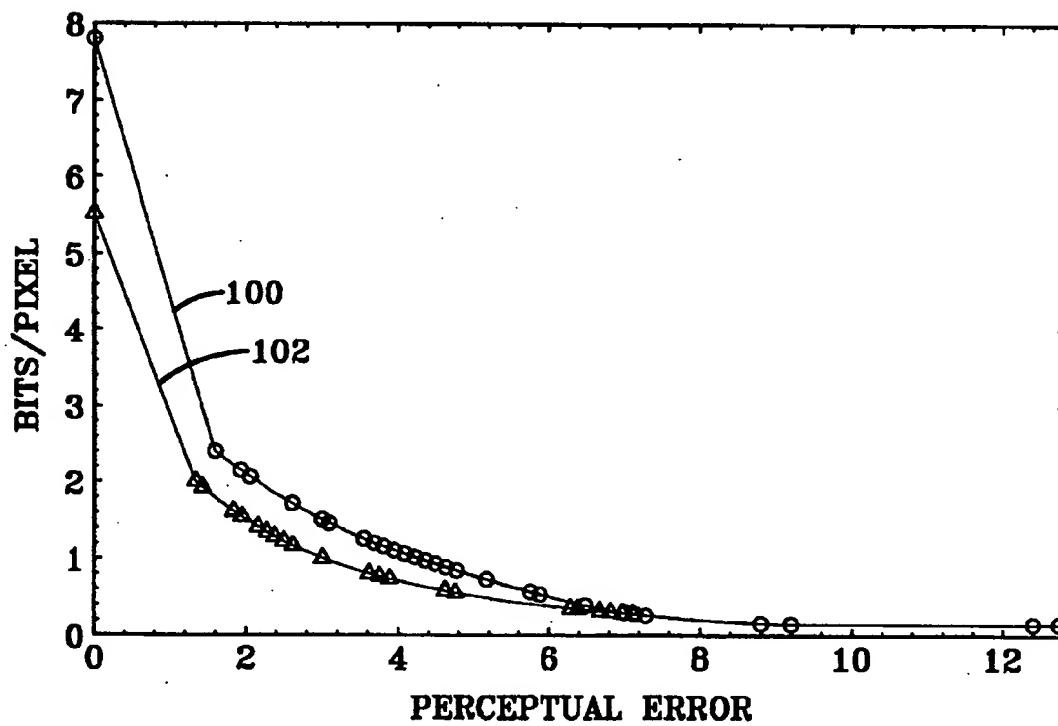


FIG-7



**FIG-8(A)**  
**PRIOR ART**



**FIG-8(B)**



# IMAGE DATA COMPRESSION HAVING MINIMUM PERCEPTUAL ERROR

## ORIGIN OF THE DISCLOSURE

The invention described herein was made by an employee of the National Aeronautics and Space Administration and it may be manufactured and used by and for the United States Government for governmental purposes without the payment of royalties thereon or therefore.

## BACKGROUND OF THE INVENTION

### A. Technical Field of Field of the Invention:

The present invention relates to an apparatus and method for coding images, and more particularly, to an apparatus and method for compressing images to a reduced number of bits by employing a Discrete Cosine Transform (DCT) in combination with visual masking including luminance and contrast techniques as well as error pooling techniques all to yield a quantization matrix optimizer that provides an image having a minimum perceptual error for a given bit rate, or a minimum bit rate for a given perceptual error.

### B. Description of the Prior Art:

Considerable research has been conducted in the field of data compression, especially the compression of digital information of digital images. Digital images comprise a rapidly growing segment of the digital information stored and communicated by science, commerce, industry and government. Digital images transmission has gained significant importance in highly advanced television systems, such as high definition television using digital information. Because a relatively large number of digital bits are required to represent digital images, a difficult burden is placed on the infrastructure of the computer communication networks involved with the creation, transmission and re-creation of digital images. For this reason, there is a need to compress digital images to a smaller number of bits, by reducing redundancy and invisible image components of the images themselves.

A system that performs image compression is disclosed in U.S. Pat. No. 5,121,216 of C. H. Chen et al, issued Jun. 9, 1992, and herein incorporated reference. The '216 patent discloses a transform coding algorithm for a still image, wherein the image is divided into small blocks of pixels. For example, each block of pixels may be either an 8x8 or 16x16 block. Each block of pixels then undergoes a two dimensional transform to produce a two dimensional array of transform coefficients. For still image coding applications, a Discrete Cosine Transform (DCT) is utilized to provide the orthogonal transform.

In addition to the '216 patent, the Discrete Cosine Transform is also employed in a number of current and future international standards, concerned with digital image compression, commonly referred to as JPEG and MPEG, which are acronyms for Joint Photographic Experts Group and Moving Pictures Experts Group, respectively. After a block of pixels of the '216 patent undergoes a Discrete Cosine Transform (DCT), the resulting transform coefficients are subject to compression by thresholding and quantization operations. Thresholding involves setting all coefficients whose magnitude is smaller than a threshold value equal to zero, whereas quantization involves scaling a coefficient by step size and rounding off to the nearest integer.

Commonly, the quantization of each DCT coefficient is determined by an entry in a quantization matrix. It is this matrix that is primarily responsible for the perceived image

quality and the bit rate of the transmission of the image. The perceived image quality is important because the human visual system can tolerate a certain amount of degradation of an image without being alerted to a noticeable error. Therefore, certain images can be transmitted at a low bit rate, whereas other images cannot tolerate any degradation and should be transmitted at a higher bit rate in order to preserve their informational content.

The '216 patent discloses a method for the compression of image information based on human visual sensitivity to quantization errors. In the method of '216 patent, there is a quantization characteristic associated with block to block components of an image. This quantization characteristic is based on a busyness measurement of the image. The method of '216 patent does not compute a complete quantization matrix, but rather only a single scalar quantizer.

Two other methods are available for computing DCT quantization matrices based on human sensitivity. One is based on a mathematical formula for human contrast sensitivity function, scaled for viewing distance and display resolution, and is disclosed in U.S. Pat. No. 4,780,761 of S. J. Daly et al. The second is based on a formula for the visibility of individual DCT basic functions, as a function of viewing distance, display resolution, and display luminance. The second formula is disclosed in a first technical article entitled "Luminance-Model-Based DCT Quantization For Color Image Compression" of A. J. Ahumada and H. A. Peterson published in 1992 in the *Human Vision, Visual Processing, and Digital Display III* Proc. SPIE 1666, Paper 32, and a second technical article entitled "An Improved Detection Model for DCT Coefficient Quantization" of H. A. Peterson, et al., published in 1993, in *Human Vision, Visual Processing and Digital Display VI* Proc. SPIE. Vol. 1913 pages 191-201 and a third technical article entitled "A visual detection model for DCT coefficient quantization" A. J. Ahumada, Jr. and H. A. Peterson, published in 1993, in *Computing in Aerospace 9*, American Institute of Aeronautics and Astronautics, pages 314-318.

The methods described in the '761 patent and the three technical articles do not adapt the quantization matrix to the image being compressed, and do not therefore take advantage of masking techniques for quantization errors that utilize the image itself. Each of these techniques has features and benefits described below.

First, visual thresholds increase with background luminance and this feature should be advantageously utilized. However, the formula given in the both referenced technical articles describes the threshold for DCT basic functions as a function of mean luminance. This would normally be taken as the mean luminance of the display. However, variations in local mean luminance within the image will in fact produce substantial variations in the DCT threshold quantities. These variations are referred to herein as "luminance masking" and should be fully taken into account.

Second, threshold for a visual pattern is typically reduced in the presence of other patterns, particularly those of similar spatial frequency and orientation. This reduction phenomenon is usually called "contrast masking." This means that a threshold error in a particular DCT coefficient in a particular block of the image will be a function of the value of that coefficient in the original image. The knowledge of this function should be taken advantage of in order to compress the image while not reducing the quality of the compressed image.

Third, the method disclosed in the two referenced technical articles ensures that a single error is below a prede-

terminated threshold. However, in a typical image there are many errors of varying magnitudes that are not properly handled by a single threshold quantity. The visibility of this error ensemble selected to handle all varying magnitudes is not generally equal to the visibility of the largest error, but rather reflects a pooling of errors over both frequencies and blocks of the image. This pooling is herein term "error pooling" and is beneficial in compressing the digital information of the image while not degrading the quality of the image.

Fourth, when all errors are kept below a perceptual threshold, a certain bit rate will result, but at times it may be desired to have an even lower bit rate. The two referenced technical articles do not disclose any method that would yield a minimum perceptual error for a given bit rate, or a minimum bit rate for a given perceptual error. It is desired that such a method be provided to accommodate this need.

Fifth, since color images comprise a great proportion of images in common use, it is desirable that the above advantages be applied to both grayscale and color images. The referenced technical articles provide a method for computing three quantization matrices for the three color channels of a color images, but do not disclose any method for optimizing the matrix for a particular color image.

Finally, it is desired that all of the above prior art limitations and drawbacks be eliminated so that a digital image may be represented by a reduced number of digital bits while at the same time providing an image having a low perceptual error.

Accordingly, an object of the present invention is to provide a method to compress digital information yet provide a visually optimized image.

Another object of the present invention is to provide a method of compressing a visual image based on luminance masking, contrast masking, and error pooling techniques.

A further object of the present invention is to provide a quantization matrix that is adapted to the individual image being compressed so that either the grayscale or the color image that is reproduced has a minimal perceptual error for a given bit rate, or a minimum bit rate for a given perceptual error.

### SUMMARY OF THE INVENTION

The invention is directed to digital compression of color or grayscale images, comprising a plurality of color channels and a plurality of blocks of pixels, that uses the DCT transform coefficients yielded from a Discrete Cosine Transform (DCT) of all the blocks as well as other display and perceptual parameters all to generate quantization matrices which, in turn, yield a reproduced image having a low perceptual error for a given bit rate. The invention adapts or customizes the quantization matrices to the image being compressed.

The present invention transforms a digital grayscale or color image into a compressed digital representation of that image and comprises the steps of transforming each color pixel, if necessary, into brightness and color channel values, down-sampling, if necessary, each color channel, partitioning each color channel into square blocks of contiguous pixels, applying a Discrete Cosine Transform (DCT) to each block in each color channel, selecting a DCT mask ( $m_{u,v,\theta}$ ) for each block of pixels in each color channel, and selecting a quantization matrix ( $q_{u,v,\theta}$ ) for quantizing DCT transformation coefficients ( $c_{u,v,\theta}$ ) produced by the DCT transformation. The application of a Discrete Cosine Transform (DCT) transforms the block of pixels into a digital signal

represented by the DCT coefficients ( $c_{u,v,\theta}$ ). The DCT mask is based on parameters comprising DCT coefficients ( $c_{u,v,\theta}$ ), and display and perceptual parameters. The selection of the quantization matrix ( $q_{u,v,\theta}$ ) comprises the steps of: (i) selecting an initial value of  $q_{u,v,\theta}$ ; (ii) quantizing the DCT coefficient  $c_{u,v,\theta}$  in each block  $b$  to form quantized coefficient  $k_{u,v,\theta}$ ; (iii) inverse quantizing  $k_{u,v,\theta}$  by multiplying by  $q_{u,v,\theta}$ ; (iv) subtracting the reconstructed coefficient  $q_{u,v,\theta}k_{u,v,\theta}$  from  $c_{u,v,\theta}$  to compute the quantization error  $e_{u,v,\theta}$ ; (v) dividing  $e_{u,v,\theta}$  by the DCT mask  $m_{u,v,\theta}$  to obtain perceptual error; (vi) pooling the perceptual errors of one frequency  $u,v$  and color  $\theta$  over all blocks  $b$  to obtain an entry in the perceptual error matrices  $p_{u,v,\theta}$ ; (vii) repeating (i-vi) for each frequency  $u,v$  and color  $\theta$ ; (viii) adjusting the values of  $q_{u,v,\theta}$  up or down until each entry in the perceptual error matrices  $p_{u,v,\theta}$  is within a target range, and entropy coding the quantization matrices and the quantized coefficients of the image.

### BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a block diagram of a computer network that may be used in the practice of the present invention.

FIG. 2 schematically illustrates some of the steps involved with the method of the present invention.

FIG. 3 schematically illustrates the steps involved, in one embodiment, with the formation of the quantization matrix optimizer of the present invention.

FIG. 4 schematically illustrates the steps involved, in another embodiment, with the formation of the quantization matrix optimizer of the present invention.

FIG. 5 illustrates a series of plots showing the variations of luminance-adjusted thresholds involved in the practice of the present invention.

FIG. 6 is a plot of the contrast masking function involved in the practice of the present invention.

FIG. 7 illustrates two plots each of a different digital image and each showing the relationship between the perceptual error and the bit rate involved in image compression in the present invention.

FIG. 8 is composed of photos A and B respectively illustrating an image compressed with and without the present invention, at equal bit rates.

### DETAILED DESCRIPTION OF THE INVENTION

Referring now to the drawings wherein like reference numerals designate like elements, there is shown in FIG. 1 a block diagram of a computer network 10 that may be used in the practice of the present invention. The network 10 is particularly suited for performing the method of the present invention related to images that may be stored, retrieved or transmitted. For the embodiment shown in FIG. 1, a first group of channelized equipment 12 and a second group of channelized equipment 14 are provided. Further, as will be further described, for the embodiment shown in FIG. 1, the channelized equipment 12 is used to perform the storage mode 16/retrieval mode 18 operations of the network 10 and, similarly, the channelized equipment 14 is used to perform the storage mode 16/retrieval mode 18 operations of the network 10. As will be further described, the storage mode 16 is shown as accessing each disk subsystem 20, whereas the retrieval mode 18 is shown as recovering information from each disk subsystem 20. Each of the channelized equipments 12 and 14 may be a SUN SPARC computer station whose operation is disclosed in instruction

manual Sun Microsystems Part #800-5701-10. Each of the channelized equipments 12 and 14 is comprised of elements having the reference numbers given in Table 1.

TABLE 1

Reference No.	Element
20	Disk Subsystem
22	Communication Channel
24	CPU Processor
26	Random Access Memory (RAM)
28	Display Subsystem

In general, and as to be more fully described, the method of the present invention, being run in the network 10, utilizes, in part, a Discrete Cosine Transform (DCT), discussed in the "Background" section, to accomplish image compression. In the storage mode 16, an original image 30, represented by a plurality of digital bits, is received from a scanner or other source at the communication channel 22 of the channelized equipment 12. The image 30 is treated as a digital file containing pixel data. The channelized equipment 12, in particular the CPU processor 24, performs a DCT transformation, computes a DCT mask and iteratively estimates a quantization matrix optimizer. The channelized equipment 12 then quantizes the digital bits comprising the image 30, and performs run-length encoding and Huffman or arithmetic coding of the quantized DCT coefficients. Run-length encoding, arithmetic coding and Huffman coding are well-known and reference may be respectively made to the discussion of reference numbers 24 and 28 of U.S. Pat. No. 5,170,264, herein incorporated by reference, for a further discussion thereof. The optimized quantization matrix is then stored in coded form along with coded coefficient data, following a JPEG or other standard. The compressed file is then stored on the disk subsystem 20 of the channelized equipment 12.

In the retrieval mode 18, the channelized equipment 12 (or 14) retrieves the compressed file from the disk subsystem 20, and decodes the quantization matrix and the DCT coefficient data. The channelized equipment 12, or the channelized equipment 14 to be described, then de-quantizes the coefficients by multiplication of the quantization matrix and performs an inverse DCT. The resulting digital file containing pixel data is available for display on the display subsystem 28 of the channelized equipment 12 or can be transmitted to the channelized equipment 14 or elsewhere by the communication channel 22. The resulting digital file is shown in FIG. 1 as 30' (IMAGE). The operation of the present invention may be further described with reference to FIG. 2.

FIG. 2 is primarily segmented to illustrate the storage mode 16 and the retrieval mode 18. FIG. 2 illustrates that the storage mode 16 is accomplished in channelized equipment, such as channelized equipment 12, and the retrieval mode is accomplished in the same or another channelized equipment, such as channelized equipment 14. The channelized equipments 12 and 14 are interfaced to each other by the communication channel 22. The image 30 being compressed by the operation of the present invention comprises a two-dimensional array of pixels, e.g., 256x256 pixels. In the case of a color image, each pixel is represented by three numbers, such as red, green, and blue components (RGB); in the case of a grayscale image, each pixel is represented by a single number representing the brightness. This array of pixels is composed of contiguous blocks; e.g., 8x8 blocks, of pixels representatively shown in segment 33. The storage

mode 16 is segmented into the following steps: color transform 31, down-sample 32, block 33, DCT 34, initial matrices, quantization matrix optimizer 36, quantize 38, and entropy code 40. The retrieval mode 18 is segmented into the following steps: entropy decode 42, de-quantize 44, inverse DCT 46, un-block 47, up-sample 48 and inverse color transform 49. The steps shown in FIG. 2 (to be further discussed with reference to FIG. 3) are associated with the image compression of the present invention and, in order to more clearly describe such compression, reference is first made to the quantities listed in the Table 2 having a general definition given therein.

TABLE 2

Quantities	General Definition
u,v	indexes of the DCT frequency (or basis function)
b	index of a block of the image
$\theta$	index of quantization color space
$c_{u,v,\theta}$	DCT coefficients of an image block
$q_{u,v,\theta}$	quantization matrix
$k_{u,v,\theta}$	quantized DCT coefficients
$e_{u,v,\theta}$	DCT error
$t_{u,v,\theta}$	threshold matrices (based on global mean luminance)
$V[u,v,\theta, Y_p, Y_q, \dots]$	threshold formula of Peterson et al. given in the article "An Improved Detection Model for DCT Coefficient Quantization" (previously cited) or similar formula
$a_{u,v,\theta}$	luminance-adjusted threshold matrices
$\alpha$	luminance masking exponent
$w_{u,v,\theta}$	contrast masking exponent (Weber exponent)
$m_{u,v,\theta}$	DCT Mask (threshold matrices adjusted for local luminance and contrast)
$j_{u,v,\theta}$	perceptual error in a particular frequency u, v, block b, color $\theta$
$p_{u,v,\theta}$	perceptual error matrix
$\beta$	spatial error-pooling exponent
$c_{0,0,b,Y}$	DC coefficient in brightness channel in block b
Y	mean luminance of the display
$\bar{c}_{0,0,Y}$	Average brightness channel DC coefficient (typically 1024)
$\Psi$	target total perceptual error value

Each pixel (step 31) is transformed from the original color representation (such as RGB) to a color representation consisting of one brightness signal and two color signals (such as the well known color space YCbCr). This new color space will be called the quantization color space and its three channels indexed by  $\theta$ . If the image is already represented in a space like YCbCr, or if the image is grayscale, then this color transformation step is skipped.

The three component color images (for example Y, Cb, and Cr) are then individually down-sampled by some factors. Down-sampling is well known and may be such as described in the technical article entitled "The JPEG still picture compression standard," by G. Wallace, published in 1991 in *Communications of the ACM*, volume 34, pages 30-44. Typically, Y is not down-sampled, and Cb and Cr are each down-sampled by a factor of two in both horizontal and vertical dimensions. If the image is grayscale, this down-sampling step is skipped.

Each down-sampled image in each color channel is then partitioned into blocks of contiguous (typically) 8x8 pixels (step 33). Each block of pixels in each color channel is subjected to the application of a Discrete Cosine Transform (DCT) (step 34) yielding related DCT coefficients. The two-dimensional Discrete Cosine Transform (DCT) is well known and may be such as described in the previously

incorporated by reference U.S. Pat. No. 5,121,216. The coefficients of the DCT, herein termed  $c_{u,v,b,\theta}$ , obtained by the Discrete Cosine Transform (DCT) of each block of pixels comprise DC and AC components. The DC coefficient in the brightness channel is herein termed  $c_{0,0,b,Y}$  which represents the average brightness of the block. The remainder of the coefficients  $c_{u,v,b,\theta}$  are termed AC coefficients.

The DCT (step 34) of all blocks (step 33), along with the display and perceptual parameters (to be described) and initial matrices, are all inputted into a quantization matrix optimizer 36, which is a process that creates an optimized quantization matrix which is used to quantize (step 38) the DCT coefficients. The optimized quantization matrix is also transferred, by the communication channel 22 of the channelized equipment 12, for its use in the retrieval mode 18 that is accomplished in the channelized equipment 14. The quantized DCT coefficients ( $k_{u,v,b,\theta}$ ) are entropy coded (step 40) and then sent to the communication channel 22. Entropy coding is well-known in the communication art and is a technique wherein the amount of information in a message is based on  $\log_n$ , where  $n$  is the number of possible equivalent messages contained in such information.

At the receiving channelized equipment 14, an inverse process occurs to reconstruct the original block of pixels thus, the received bit stream of digital information containing quantized DCT coefficients  $k_{u,v,b,\theta}$  is entropy decoded (step 42) and then are de-quantized (step 44), such as by multiplying by the quantization step size  $q_{u,v,\theta}$  to be described. An inverse transform (step 46), such as an inverse Discrete Cosine Transform (DCT), is then applied to the DCT coefficients ( $c'_{u,v,b,\theta}$ ) to reconstruct the block of pixels. After the reconstruction, the block of pixels are unblocked, up-sampled, and inverse color transformed so as to provide a reconstituted and reconstructed image 30'. The quantization matrix optimizer 36 is of particular importance to the present invention and may be described with reference to FIG. 3.

The quantization optimizer matrix 36 is adapted to the particular image being compressed and, as will be further described, advantageously includes the functions of luminance masking, contrast masking, error pooling and selectable quality. All of which functions cooperate to yield a compressed image having a minimal perceptual error for a given bit rate, or minimum bit rate for a given perceptual error. The quantization matrix optimizer 36, in one embodiment, comprises a plurality of processing segments each having a reference number and nomenclature given in Table 3.

TABLE 3

Processing Segment	Nomenclature
50	Compute visual thresholds
52	Adjust thresholds in each block for block mean brightness
54	Adjust thresholds in each block for block component contrast
56	Quantize
58	Compute quantization error
60	Scale error by DCT mask
62	Pool error over blocks
64	Pooled error matrix - target error?
66	Adjust quantization matrices

The first step in the generation of the quantization matrix optimizer 36 is the derivation of a function DCT mask 70 which is accomplished by the operation of processing segments 50, 52 and 54 and is determined, in part, by the

display and perceptual parameters 72 having typical values given in the below Table 4.

TABLE 4

Display and Perceptual Parameters	Typical Values
$a_l$	0.65
$\beta$	4
$w_{u,v,\theta}$	0.7
$Y$	40 cd/m <sup>2</sup>
image grey levels	256
$c_{0,0,b,Y}$	1024
$r_Y$	0.05 (veiling luminance, expressed as ratio of display mean luminance $Y$ )
$P_H, P_V$	32, 32 (these define the number of pixels per degree of visual angle in horizontal and vertical directions. These values correspond to a 256 x 256 pixel image at a viewing distance of 7.125 picture heights)

The display and perceptual parameters 72 are used to compute a matrix of DCT component visual thresholds by using a formula such as that more fully described in the previously referenced first, second, and third technical articles and which formula may be represented by expression 1:

$$t_{u,v,\theta} = V[u, v, \theta, Y, P_H, P_V, \dots]$$

where  $V$  represents the threshold formula of Table 2,  $u$  and  $v$  are indexes of the DCT frequency,  $\theta$  is the quantization color space,  $Y$  is the mean luminance of the display,  $P_H$  represents pixels per degree of visual angle horizontal and  $P_V$  represents pixels per degree of visual angle vertical.

The visual threshold values of expression (1) are then adjusted for mean block luminance in processing segment 52. The processing segment 52 receives only the DC coefficient of the DCT coefficients indicated by reference number 74, whereas segment 54 receives and uses the entire DCT coefficients. The formula used to accomplish processing segment 52 is given by expression 2:

$$a_{u,v,b,\theta} = t_{u,v,\theta} \left( \frac{r_Y + c_{0,0,b,Y} \sqrt{c_{0,0,b,Y}}}{r_Y + 1} \right)^{\alpha}$$

where  $a_l$  is a luminance-masking exponent having a typical value of 0.65,  $a_{u,v,b,\theta}$  is the adjusted threshold,  $t_{u,v,\theta}$  is the un-adjusted threshold,  $c_{0,0,b,Y}$  is the average of the DC terms of the DC coefficients for the present image, or may be simply a nominal value of 1024, for an eight (8) bit image, and  $c_{0,0,b,Y}$  is the DC term of the DCT for block  $b$  in the brightness channel. The term  $r_Y$  is the veiling luminance (luminance cast on the display by room lights etc.), expressed as ratio of display mean luminance  $Y$ . A typical value of  $r_Y$  is 0.05.

As seen in FIG. 3, the luminance-adjusted thresholds of segment 52 are then adjusted for component contrast by the operation of a routine having a relationship as given by the below expression 3:

$$m_{u,v,b,\theta} = a_{u,v,b,\theta} \text{Max} \left[ 1, \left| \frac{c_{u,v,b,\theta}}{a_{u,v,b,\theta}} \right|^{w_{u,v,\theta}} \right]$$

where  $m_{u,v,b,\theta}$  is the contrast-adjusted threshold,  $c_{u,v,b,\theta}$  is the DCT coefficient,  $a_{u,v,b,\theta}$  is the corresponding threshold of expression 2, and  $w_{u,v,\theta}$  is the exponent that lies between 0 and 1 and typically has a value of 0.7. Because the exponent

$w_{u,v,\theta}$  may differ for different colors and frequencies of the DCT coefficients, a matrix of exponents equal in size to the quantization matrices is provided. The result of the operations of processing segments 50, 52, and 54 is the derivation of the quantity  $m_{u,v,\theta}$  herein termed "DCT mask" 70 which is supplied to the processing segment 60 to be described hereinafter.

After the calculation of the DCT mask 70 has been determined, an iterative process of estimating the quantization matrix optimizer 36 begins and is comprised of processing segments 56, 58, 60, 62, 64, and 66. The initial matrices 35, which are typically fixed and which may be any proper quantization matrices, are typically set to a maximum permissible quantization matrix entry (e.g., in the JPEG standard this maximum value is equal to 255) and are used in the quantization of the image as indicated in processing segment 56.

Each transformed block of the image contained in the initial matrix 35 is then quantized in segments 56 by dividing it, coefficient by coefficient, by the quantization matrix ( $q_{u,v,\theta}$ ), and is rounded to the nearest integer as shown in expression 4:

$$k_{u,v,\theta} = \text{Round}[c_{u,v,\theta}/q_{u,v,\theta}]$$

Segment 58 then computes the quantization error  $e_{u,v,\theta}$  in the DCT domain, which is equal to the difference between the de-quantized and original DGT coefficients  $c_{u,v,\theta}$ , and is shown by expression 5:

$$e_{u,v,\theta} = c_{u,v,\theta} - k_{u,v,\theta} q_{u,v,\theta}$$

From expressions 4 and 5, it may be shown that the maximum possible quantization error  $e_{u,v,\theta}$  is  $q_{u,v,\theta}/2$ .

The output of segment 58 is then applied to segment 60, wherein the quantization error is scaled (divided) by the value of the DCT mask 70. This scaling is described by expression 6:

$$j_{u,v,\theta} = e_{u,v,\theta}/m_{u,v,\theta}$$

where  $j_{u,v,\theta}$  is defined as the perceptual error at frequency  $u,v$  and color  $\theta$  in block  $b$ . The scaled quantization error is then applied to the processing segment 62. The processing segment 62 causes all the scaled errors to be pooled over all of the blocks, separately for each DCT frequency and color ( $u,v,\theta$ ). The term "error pooling" is meant to represent that the errors are combined over all of the blocks rather than having one relatively large error in one block dominating the other errors in the remaining blocks. The pooling is accomplished by a routine having a relationship of expression 7:

$$p_{u,v,\theta} = \left( \sum_b j_{u,v,\theta}^\beta \right)^{1/\beta}$$

Where  $j_{u,v,\theta}$  is a perceptual error in a particular frequency  $u,v$ , color  $\theta$ , and block  $b$ ,  $\beta$  is a pooling exponent having a typical value of 4. It is allowed that the routine of expression (7) provide a matrix of exponents  $\beta$  since the pooling of errors may vary for different DCT coefficients. If the three color images have been down-sampled (step 32) by different factors, then the range of the block index  $b$  will differ for the three color channels.

The matrices  $p_{u,v,\theta}$  of expression (7) are the "perceptual error matrices" and are a simple measure of the visibility of artifacts within each of the frequency bands and colors defined by the DCT basic functions. More particularly, the perceptual error matrix is a good indication of whether or not

the human eye can perceive a dilution of the image that is being compressed. The perceptual error matrices  $p_{u,v,\theta}$  developed by segments 56, 58, 60 and, finally, segment 62 are applied to processing segment 64.

In processing segment 64, each element of the perceptual error matrices  $p_{u,v,\theta}$  is compared to a target error parameter  $\psi$ , which specifies a global perceptual quality of the compressed image. This global quality is somewhat like the entries in the perceptual error matrices and again is a good indication of whether the degradation of the compressed image will be perceived by the human eye. If all quantities or errors generated by segment 62 and entered into segment 64 are within a delta of  $\psi$ , or if the errors of segment 62 are less than the target error parameter  $\psi$  and the corresponding quantization matrix entry is at a maximum (processing segment 56), the search is terminated and the current element of quantization matrix is outputted to comprise an element of the final quantization matrices 78. Otherwise, if the element of the perceptual error matrices is less than the target parameter  $\psi$ , the corresponding entry (segment 56) of the quantization matrix is incremented. Conversely, if the element of the perceptual error matrix is greater than the target parameter  $\psi$ , the corresponding entry (segment 56) of the quantization matrix is decremented. The incrementing and decrementing is accomplished by processing segment 66.

A bisection method, performed in segment 66, is typically used to determine whether to increment or decrement the initial matrices 35 entered into step 56. In the bisection method a range is established for  $q_{u,v,\theta}$  between lower and upper bounds, typically 1 and 255 respectively. The perceptual error matrix  $p_{u,v,\theta}$  is evaluated at the mid-point of the range. If  $p_{u,v,\theta}$  is greater than the target error parameter  $\psi$ , then the lower bound is reset to the mid-point, otherwise the upper bound is reset to the mid-point. This procedure is repeated until the mid-point no longer changes. As a practical matter, since the quantization matrix entries  $q_{u,v,\theta}$  in the baseline JPEG standard are eight bit integers, the needed degree of accuracy is normally obtained in nine iterations from a starting range of 1-255 (initial entry into segment 56). The output of the program segment 66 is applied to the quantize segment 56 and then steps 56-66 are repeated, if necessary, for the remaining elements in the initial matrices.

The preceding methods have been described with respect to a color image, consisting of three color channels, indexed by  $\theta$ . If compression of a grayscale image is desired, then only one color channel exists (brightness, or  $Y$ ), but otherwise all operations remain the same.

The processing segments shown in FIG. 3 yield a compressed image with a resulting bit rate; however, if a particular bit rate is desired for the image, then the processing segments shown in FIG. 4 and given in the below Table 5 are applicable.

TABLE 5

Processing Segment	Nomenclature
80	Select desired bit rate
82	Set initial target perceptual error
84	Optimize quantization matrix (56, 58, 60, 62, 64 and 66)
86	Quantize
88	Entropy code
90	Decision box (Is the bit rate = desired bit rate)
92	Adjust target perceptual error

The processing segments 86-92, shown in FIG. 4 and given in Table 5, allow for the attainment of a particular bit

rate and utilizes a second, higher-order optimization which if the first optimization results in a bit rate which is greater than desired, the value of the target perceptual error parameter  $\psi$  of segment 92 is incremented. Conversely, if a bit rate results which is lower than desired, the value of the target perceptual error parameter  $\psi$  of segment 92 is decremented.

The sequence of FIG. 4 starts with the selection (segment 80) of the desired bit rate followed by the setting (segment 82) of the initial target perceptual error. The output of segment 82, as well as the output of segment 92, is applied to segment 84 which comprises segments 56, 58, 60, 62, 64 and 66, all previously described with reference to FIG. 3 and all of which contribute to provide an optimized quantization matrix in a manner also described with reference to FIG. 3. The output of segment 84 is applied to the quantize segment 86 which operates in a similar manner as described for the quantize segment 38 of FIG. 2. The output of segment 86 is applied to the entropy code segment 88 which operates in a similar manner as described for the entropy code 40 of FIG. 2.

To accomplish the adjustment of the bit rate, the output processing segment 88 is applied to a decision segment 90 in which the actual bit rate is compared against the desired bit rate and, the result of such comparison, determines the described incrementing or decrementing of the target perceptual error parameter  $\psi$ . After such incrementing or decrementing the processing steps 84-90 is repeated until the actual bit rate is equal to the desired bit rate, and the final quantization matrix 78 is created.

It should now be appreciated that the practice of the present invention provides for a quantization matrix 78 that yields minimum perceptual error for a given bit or a minimum bit rate for a given perceptual error. The present invention, as already discussed, provides for visual masking by luminance and contrast techniques as well as by error pooling. The luminance masking feature may be further described with reference to FIG. 5.

FIG. 5 has a y axis given in the log function of  $a_{u,v,b,\theta}$  of expression (2) and a x axis given in block luminance measured in  $\text{cd/m}^2$ . The quantity  $a_{u,v,b,\theta}$  shown in FIG. 5 is based on a maximum display luminance  $L$  of  $80 \text{ cd/m}^2$ , the brightness color channel, a veiling luminance ratio  $r_v$  of 0.05, and a grey scale (reference scale for use in black-and-white television, consisting of several defined levels of brightness with neutral color) resolution of eight (8) bits.

The curves shown in FIG. 5 are plots for the DCT coefficient frequencies given Table 6.

TABLE 6

Plots	Frequency
94A	7, 7
94B	0, 7
94C	0, 3
94D	0, 1

Detection threshold for a luminance pattern typically depends upon mean luminance from the local image region. More particularly, the higher the background of the image being displayed, the higher, the luminance threshold. This is usually called "light adaptation" but it is called herein "luminance masking."

FIG. 5 illustrates this effect whereby higher background luminance yields higher luminance thresholds. The plots 94A-94D illustrate that almost one log unit change in  $a_{u,v,b,\theta}$  might be expected to occur within an image, due to variations in the mean luminance between blocks. The present

invention takes this variation into account, whereas known prior art techniques fail to consider this wide variation.

In practice, the initial calculation of  $t_{u,v,\theta}$  should be made assuming a selected displayed luminance  $Y$ . The parameter  $a_v$  has a typical value of 0.65. It should be noted, that luminance masking may be suppressed by setting  $a_v$  equal to 0. More generally,  $a_v$  controls the degree to which the masking of FIG. 4 occurs. It should be further noted that the power function given in expression 2 makes it easy to incorporate a non-unity display Gamma, by multiplying  $a_v$  by the Gamma exponent having a typical value of 2.3.

As previously discussed with reference to processing segment 54 of FIG. 3, the present invention also provides for contrast masking. Contrast masking refers to the reduction in the visibility of one image component by the presence of another. This masking is strongest when both components are of the same spatial frequency, orientation, and location within the digital image being compressed. Contrast masking is achieved in the present invention by expression (3) as previously described. The benefits of the contrast masking function is illustrated in FIG. 6.

FIG. 6 has a Y axis given in the quantity  $m_{u,v,b,\theta}$  (DCT mask) and a X axis given in the quantity  $c_{u,v,b,\theta}$  (DCT coefficient) and illustrates a response plot 98 of the DCT mask  $m_{u,v,b,\theta}$  as a function of the DCT coefficient  $c_{u,v,b,\theta}$  for the parameter  $w_{u,v,\theta}=0.7$  and  $a_{u,v,b,\theta}=2$ . Because the effect of the DC coefficient  $C_{0,0,b}$  upon the luminance masking (see FIG. 5) has already been expressed, the plot 98 does not include an effect of the DC coefficient  $c_{0,0,b}$  and accomplishes such by setting the value of  $w_{0,0,\theta}$  equal to 0. From FIG. 6 it may be seen that in this example the DCT mask ( $m_{u,v,b,\theta}$ ) increases by over a factor of three as  $c_{u,v,b,\theta}$  varies between about 2 to 10. This DCT mask ( $m_{u,v,b,\theta}$ ) generated by processing segment 54 adjusts each block for component contrast and is used in processing segment 60 to scale (divide) the quantization error with both functions of the DCT mask ensuring good digital compression, while still providing an image having good visual aspects.

It should be appreciated that the practice of the invention provides contrast masking so as to provide for a high quality visual representation of compressed digital images as compared to other prior art techniques.

The overall operation of the present invention is essentially illustrated in FIG. 7. FIG. 7 has a Y axis given in bits/pixel of the compressed digital image and a X axis given in perceptual error  $\psi$ . FIG. 7 illustrates two plots 100 and 102 for two different grayscale images that were compressed and reconstituted in accordance with the hereinbefore described principles of the present invention. From FIG. 7 it is seen that the increasing bits/pixel rate causes a decrease in the perception error.

The previously given description described herein yields desired quantization matrices  $q_{u,v,\theta}$  with a specified perceptual error  $\psi$ . However, if desired one may have a quantization matrix  $q_{u,v,\theta}$  that uses a given bit rate  $b_0$  with a minimum perceptual error  $\psi$ . This can be done iteratively by noting that the bit rate is a decreasing function of the perceptual error  $\psi$ , as shown in FIG. 7. In the practice of our present invention a second order interpolating polynomial fit to all previous estimated values of  $\{b,\psi\}$  to estimate a next candidate  $\psi$ , terminating when  $|b-b_0| < \Delta b$ , where  $\Delta b$  is the desired accuracy in bit rate. On each iteration a complete estimation of  $\psi$  is performed, as shown in FIG. 4.

An illustration of results obtained with the present invention is shown in FIG. 8. FIG. 8A shows an image compressed without the benefit of  $t$  using standard JPEG techniques. The image was  $768 \times 512$  pixels, and should be

viewed at a distance of 7 picture heights (about 20 inches) to yield 64 pixels/degree. FIG. 8B shows the same image compressed to the same bit rate using the present invention. The visual quality is clearly superior in FIG. 8B, using the present invention. The limitation of FIG. 8A is especially noted by the objectionable contouring in the sky.

It should now be appreciated that the practice of the present invention provides a perceptual error that incorporates visual masking by luminance and contrast techniques, and error pooling to estimate the matrix that has a minimum perceptual error for a given bit rate, or minimal bit rate for a given perceptual error. All told the present invention provides a digital compression technique that is useful in the transmission and reproduction of images particularly those found in high definition television applications.

Further, although the invention has been described relative to a specific embodiment thereof, it is not so limited and many modifications and variations thereof now will be readily apparent for those skilled in the art in light of the above teachings.

What I claim is:

1. A method for transforming a digital color image into a compressed representation of said image comprising the steps of:

- (a) transforming each color image pixel to three color channel values corresponding to brightness and two color signals,
- (b) down-sampling each transformed color channel by a predetermined factor,
- (c) partitioning each color channel image into a set of square blocks of pixels, each block having an index b,
- (d) applying a Discrete Cosine Transform (DCT) to transform each said block of pixels into digital signals represented by DCT transformation coefficients ( $c_{u,v,b,\theta}$ );
- (e) selecting luminance-adjusted thresholds for each block of pixels based on parameters comprising said DCT transformation coefficients  $c_{u,v,b,\theta}$  and display and perceptual parameters;
- (f) selecting a DCT mask ( $m_{u,v,b,\theta}$ ) for each block of pixels based on parameters comprising said DCT transformation coefficients  $c_{u,v,b,\theta}$  and luminance-adjusted thresholds, and said display and perceptual parameters;
- (g) selecting a quantization matrix ( $q_{u,v,\theta}$ ) comprising the steps of:
  - (i) selecting an initial value of  $q_{u,v,\theta}$ ;
  - (ii) quantizing the DCT coefficient  $c_{u,v,b,\theta}$  in each block b to form quantized coefficient  $k_{u,v,b,\theta}$ ;
  - (iii) de-quantizing the quantized coefficients by multiplying them by  $q_{u,v,\theta}$
  - (iv) subtracting the de-quantized coefficient  $q_{u,v,\theta}k_{u,v,b,\theta}$  from  $c_{u,v,b,\theta}$  to compute the quantization error  $e_{u,v,b,\theta}$ ;
  - (v) dividing  $e_{u,v,b,\theta}$  by the DCT mask  $m_{u,v,b,\theta}$  to obtain perceptual errors  $j_{u,v,b,\theta}$ ;
  - (vi) pooling the perceptual errors of one frequency u,v and color channel  $\theta$  over all blocks b to obtain an entry in the perceptual error matrices  $p_{u,v,\theta}$ ;
  - (vii) repeating steps i-vi for each frequency u,v; $\theta$ ;
  - (viii) adjusting the values  $q_{u,v,\theta}$  up or down until each entry in the perceptual error matrix  $p_{u,v,\theta}$  is within a target range
- (h) entropy coding said quantization matrices and the quantized coefficients of said image.

2. The method of transforming an image according to claim 1 further comprising the steps:

- (a) transmitting entropy coded image to a user of said digital image;
- (b) decoding said entropy coded image;
- (c) decoding said quantization matrix to derive said DCT transformation coefficients;
- (d) applying an inverse Discrete Cosine Transform to derive said block of pixels;
- (e) up-sampling each color channel image; and
- (f) inverse color transforming each color pixel to recover the reconstructed color image.

3. The method of transforming an image according to claim 1, wherein luminance-adjusted thresholds are determined by the following expression:

$$a_{u,v,b,\theta} = t_{u,v,\theta} \left( \frac{r_Y + c_{0,0,b,Y} \bar{c}_{0,0,Y}}{1 + r_Y} \right)^a$$

where  $a_{u,v,b,\theta}$  are the luminance-adjusted thresholds,  $r_Y$  is a veiling luminance quantity,  $a$  is a luminance-masking exponent having a value of about 0.65,  $t_{u,v,\theta}$  are the un-adjusted thresholds,  $c_{0,0,Y}$  is an average of the DC terms of the DCT coefficients  $c_{u,v,b,\theta}$  of the image, and  $c_{0,0,b,Y}$  is the DC coefficient of block b of a brightness (Y) channel.

4. The method of transforming an image according to claim 1, wherein said DCT mask is determined by the following expression:

$$m_{u,v,b,\theta} = a_{u,v,b,\theta} \text{Max} \left[ 1, \left| \frac{c_{u,v,b,\theta}}{a_{u,v,b,\theta}} \right|^{w_{u,v,\theta}} \right]$$

where  $c_{u,v,b,\theta}$  is the DCT coefficient,  $a_{u,v,b,\theta}$  is a corresponding block luminance-adjusted threshold, and  $w_{u,v,\theta}$  is an exponent that lies between 0 and 1 and is typically about 0.7, and where  $w_{0,0,\theta} = 0$ .

5. The method of transforming an image according to claim 1, wherein said pooling of perceptual errors is determined by the following expression:

$$p_{u,v,\theta} = \left( \sum_b j_{u,v,b,\theta}^\beta \right)^{1/\beta}$$

where  $j_{u,v,b,\theta}$  is a perceptual error in a particular frequency u,v, color  $\theta$  and block b and  $\beta$  is a pooling exponent having a typical value of about 4.

6. A computer involved in converting a digital color image into a compressed representation of said image, said computer comprising:

- (a) means for transforming each color pixel into three color channel values corresponding to brightness and two color channels;
- (b) means for down-sampling each color channel;
- (c) means for partitioning each color channel into a set of square blocks of pixels, each block having an index b,
- (d) means for applying a Discrete Cosine Transform (DCT) to transform said blocks of pixels into digital signals represented by DCT transformation coefficients ( $c_{u,v,b,\theta}$ );
- (e) means for selecting a DCT mask ( $m_{u,v,b,\theta}$ ) for each block of pixels based on parameters comprising said DCT transformation coefficients  $c_{u,v,b,\theta}$  and display and perceptual parameters;
- (f) means for selecting a quantization matrix comprising:
  - (i) means for selecting an initial value of  $q_{u,v,\theta}$ ;
  - (ii) means for quantizing the DCT coefficient  $c_{u,v,b,\theta}$  in each block b to form quantized coefficient  $k_{u,v,b,\theta}$ ;



- (iii) means for de-quantizing the quantized coefficients by multiplying them by  $q_{u,v,\theta}$ ;
  - (iv) means for subtracting the de-quantized coefficient  $q_{u,v,\theta}k_{u,v,b,\theta}$  from  $c_{u,v,b,\theta}$  to compute a quantization error  $e_{u,v,b,\theta}$ ;
  - (v) means for dividing  $e_{u,v,b,\theta}$  by the DCT mask  $m_{u,v,b,\theta}$  to obtain perceptual errors;
  - (vi) means for pooling the perceptual errors of one frequency  $u,v$  and color  $\theta$  over all blocks  $b$  to obtain an entry in the perceptual error matrix  $p_{u,v,\theta}$ ;
  - (vii) means for adjusting the values  $q_{u,v,\theta}$  up or down until each entry in the perceptual error matrix is within a target range.
  - (g) means for entropy coding said quantization matrix and the quantized coefficients of said image.
7. The computer involved in converting an image according to claim 6 further comprising:
- (a) means for transmitting entropy coded image to a user of said digital image;
  - (b) means for decoding said entropy coded image;
  - (c) means for decoding said quantization matrix to derive said DCT transform coefficients;
  - (d) means for applying an inverse Discrete Cosine Transform to derive said block of pixels;
  - (e) means for reassembling said blocks to recover the color channels
  - (f) means for up-sampling said color channels to recover full resolution color channels; and
  - (g) means for inverse color transforming said color channels to recover the reconstructed image.
8. The computer involved in converting an image according to claim 6, wherein luminance-adjusted thresholds are determined by the following expression:

$$a_{u,v,b,\theta} = t_{u,v,\theta} \left( \frac{r\gamma + c_{0,0,b,\theta} \bar{c}_{0,0,\gamma}}{1 + r\gamma} \right)^{\gamma}$$

- 5 where  $a_{u,v,b,\theta}$  are luminance-adjusted thresholds,  $r_{\gamma}$  is a veiling luminance quantity  $a_{\gamma}$  is a luminance-masking exponent having a value of about 0.65,  $t_{u,v,\theta}$  are the un-adjusted thresholds,  $\bar{c}_{0,0,\gamma}$  is an average of the DC terms of the DCT coefficients  $c_{u,v,b,\theta}$  of the image, and  $c_{0,0,b,\gamma}$  is the DC
- 10 coefficient of block  $b$  of a brightness (Y) channel.

9. The computer involved in converting an image according to claim 6, wherein said DCT mask is determined by the following expression:

$$15 \quad m_{u,v,b,\theta} = a_{u,v,b,\theta} \text{Max} \left[ 1, \left| \frac{c_{u,v,b,\theta}}{a_{u,v,b,\theta}} \right|^{w_{u,v,\theta}} \right]$$

- 20 where  $c_{u,v,b,\theta}$  is the DCT coefficient,  $a_{u,v,b,\theta}$  is a corresponding block luminance-adjusted threshold, and  $w_{u,v,\theta}$  is an exponent that lies between 0 and 1 and is typically about 0.7, and where  $w_{0,0,\theta} = 0$ .

10. The computer involved in converting an image according to claim 6, wherein said pooling of perceptual

25 errors is determined by the following expression:

$$p_{u,v,\theta} = \left( \sum_b j_{u,v,b,\theta}^{\beta} \right)^{1/\beta}$$

- 30 where  $j_{u,v,b,\theta}$  is a perceptual error in a particular frequency  $u,v$ , color  $\theta$  and block  $b$  and  $\beta$  is a pooling exponent having a typical value of about 4.

\* \* \* \* \*